CENTRAL LIMIT THEOREM FOR RANDOM PARTITIONS UNDER THE PLANCHEREL MEASURE

L. V. Bogachev* **and Z. G. Su****

* Department of Statistics, University of Leeds, United Kingdom.
E-mail: bogachev@maths.leeds.ac.uk

** Department of Mathematics, Zhejiang University, Hangzhou, P.R. China.
E-mail: suzhonggen@zju.edu.cn

To the memory of Sergei Kerov (1946–2000)

1. INTRODUCTION

A *partition* of a natural number n is any integer sequence $\lambda = (\lambda_1, \lambda_2, \dots)$ such that $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$ and $\lambda_1 + \lambda_2 + \cdots = n$ (notation: $\lambda \vdash n$). In particular, $\lambda_1 = \max\{\lambda_i \in \lambda\}$. Every partition $\lambda \vdash n$ can be represented geometrically by a planar shape called the *Young diagram*, consisting of n unit cell arranged in consecutive columns, containing $\lambda_1, \lambda_2, \dots$ cells, respectively.

On the set $\mathcal{P}_n := \{\lambda \vdash n\}$ of all partitions of a given n, consider the *Plancherel measure*

$$P_n(\lambda) := \frac{d_\lambda^2}{n!}, \qquad \lambda \in \mathcal{P}_n, \tag{1}$$

where d_λ is the number of standard tableaux of a given shape λ, that is, the total number of arrangements of the numbers $1, \dots, n$ in the cells of the Young diagram $\lambda \in \mathcal{P}_n$, such that the numbers increase in each row (from left to right) and each column (bottom up). Note that d_λ also equals the dimension of the irreducible (complex) representation of the symmetric group $\mathfrak{S}_n$ (i.e., the group of permutations of order n), indexed by the partition λ (see [7, 18, 19]). According to the RSK[1] correspondence (see [7]), any permutation $\sigma \in \mathfrak{S}_n$ is associated with exactly one (ordered) pair of standard tableaux of the same shape $\lambda \in \mathcal{P}_n$. Since there are $n!$ such permutations, this implies the Burnside identity (see [19])

$$\sum_{\lambda \in \mathcal{P}_n} d_\lambda^2 = n!,$$

thus the measure P_n defined in (1) determines a probability distribution on $\mathcal{P}_n$.

The Plancherel measure arises naturally in representation-theoretic, combinatorial, and probabilistic problems (see [6]). For example, the RSK correspondence implies

[1] Robinson–Schensted–Knuth.

that the largest term λ_1 of the partition λ associated with a given permutation $\sigma \in \mathfrak{S}_n$ equals the length ℓ_n of the longest increasing subsequence contained in σ. Therefore, the Plancherel distribution of λ_1 coincides with the distribution of ℓ_n in a random (uniformly distributed) permutation $\sigma \in S_n$ (see [2]).[2]

The upper boundary of the Young diagram corresponding to the partition $\lambda \in \mathcal{P}_n$ can be viewed as the graph of a stepwise (left-continuous) function $\lambda(x)$, $x \geq 0$, defined by

$$\lambda(x) := \lambda_0 \mathbf{1}_{\{0\}}(x) + \sum_{i=1}^{\infty} \lambda_i \mathbf{1}_{(i-1,i]}(x) \equiv \lambda_{\lceil x \rceil}, \tag{2}$$

where $\mathbf{1}_B(x)$ is the characteristic function (indicator) of set B and $\lceil x \rceil := \min\{m \in \mathbb{Z} : m \geq x\}$ is the ceiling integer part of x. Logan and Shepp [13] and, independently, Vershik and Kerov [18] (see also [19]) have discovered that, as $n \to \infty$, a typical Young diagram, suitably scaled, has a "limit shape" represented by the graph of some function $y = \omega(x)$. This means that for the overwhelming majority of partitions $\lambda \in \mathcal{P}_n$ (with respect to the Plancherel measure P_n), the boundary of their scaled Young diagrams is contained in an arbitrarily small vicinity of the graph of $\omega(x)$.

More specifically, set

$$\bar{\lambda}_n(x) := \frac{1}{\sqrt{n}}\, \lambda(\sqrt{n}x), \qquad x \geq 0, \tag{3}$$

and consider the function $y = \omega(x)$ defined by the parametric equations

$$x = \frac{2}{\pi}\,(\sin\theta - \theta\cos\theta), \qquad y = x + 2\cos\theta, \qquad 0 \leq \theta \leq \pi. \tag{4}$$

The function $\omega(x)$ is decreasing on $[0,2]$ and $\omega(0) = 2$, $\omega(2) = 0$. Define $\omega(x)$ as zero for all $x > 2$. Then the random process $\bar{\lambda}_n(x)$ satisfies the following law of large numbers [18, 19, 13]:

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} P_n\Bigl\{\sup_{x\geq 0}\bigl|\bar{\lambda}_n(x) - \omega(x)\bigr| > \varepsilon\Bigr\} = 0.$$

In particular, for $x = 0$ it follows that the maximal term in a typical partition asymptotically behaves like $2\sqrt{n}$:

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} P_n\Bigl\{\Bigl|\frac{\lambda_1}{\sqrt{n}} - 2\Bigr| > \varepsilon\Bigr\} = 0. \tag{5}$$

Remark 1. Due to the invariance of the Plancherel measure under the transposition of Young diagrams $\lambda \leftrightarrow \lambda'$ (when the columns of the diagram λ become rows of the transposed diagram λ' and vice versa), the same law of large numbers holds for $\lambda'_1 = \#\{\lambda_i \in \lambda\}$ (i.e., for the number of terms in the random partition λ).

---

[2] The interest in the asymptotic behavior of the random variable ℓ_n was stimulated by the Ulam problem (see [6]). The problem was settled by Vershik and Kerov [18] who showed that $\ell_n/\sqrt{n} \to 2$ in probability (cf. formula (5) below).

A natural question about fluctuations of the random function $\bar{\lambda}_n$ around the limit curve ω was posed in [13] (see also [19]), but it remained open for more than 15 years. Kerov [12] (see also [9]) gave a partial answer by establishing the convergence of the random process

$$\Delta_n(x) := \sqrt{n}\,\bigl(\bar{\lambda}_n(x) - \omega(x)\bigr) \tag{6}$$

to a generalized Gaussian process (without any further normalization!). To state this result more precisely, it is convenient to pass to the coordinates $u = x - y$, $v = x + y$, which corresponds to anticlockwise rotation by 45° and dilation by $\sqrt{2}$. In the new coordinates, the boundary of the scaled Young diagram is determined by the piecewise linear (continuous) function $\widetilde{\lambda}_n(u)$ and the limit shape is given by

$$\Omega(u) := \begin{cases} \dfrac{2}{\pi}\left(u \arcsin \dfrac{u}{2} + \sqrt{4 - u^2}\right), & |u| \le 2, \\ |u|, & |u| \ge 2. \end{cases}$$

Then, according to [12, 9], the random process

$$\widetilde{\Delta}_n(u) := \sqrt{n}\,\bigl(\widetilde{\lambda}_n(u) - \Omega(u)\bigr) \tag{7}$$

converges in distribution to a generalized Gaussian process $\widetilde{\Delta}(u)$, $u \in [-2, 2]$, defined by the formal random series

$$\widetilde{\Delta}(u)|_{u=2\cos\theta} = \frac{2}{\pi} \sum_{k=2}^{\infty} \frac{X_k \sin(k\theta)}{\sqrt{k}}\,, \qquad \theta \in [0, \pi], \tag{8}$$

where $X_2, X_3, \dots$ are independent random variables with standard normal distribution $\mathcal{N}(0,1)$. Convergence of $\widetilde{\Delta}_n(\cdot)$ to $\widetilde{\Delta}(\cdot)$ is understood in the sense of generalized functions. It is convenient to choose test functions in the form of the modified Chebyshev polynomials of the second kind (see [9]), defined by the formula

$$U_k(u)|_{u=2\cos\theta} := \frac{\sin((k+1)\theta)}{\sin\theta}\,, \qquad \theta \in [0, \pi]. \tag{9}$$

Then one can show (see details in [9]) that for $k = 2, 3, \dots$,

$$\begin{aligned} \int_{\mathbb{R}} \widetilde{\Delta}_n(u)\, U_{k-1}(u)\, du &\xrightarrow{d} \int_{-2}^{2} \widetilde{\Delta}(u)\, U_{k-1}(u)\, du \\ &= 2 \int_0^{\pi} \widetilde{\Delta}(2\cos\theta) \sin(k\theta)\, d\theta = \frac{2X_k}{\sqrt{k}} \end{aligned} \tag{10}$$

(here and below, the symbol $\xrightarrow{d}$ denotes convergence in distribution).

Remark 2. Note that $U_0(u) \equiv 1$ (see (9)), in which case we have

$$\int_{\mathbb{R}} \widetilde{\Delta}_n(u)\, U_0(u)\, du = \int_{\mathbb{R}} \widetilde{\Delta}_n(u)\, du = 2 \int_0^{\infty} \Delta_n(x)\, dx = 0,$$

because the area of the scaled Young diagram equals 1, as well as the area under the graph of $y = \omega(x)$. This explains why $k \ge 2$ in (8) and (10). Also note that fluctuations outside $[-2, 2]$ are negligible, since $\int_{|u|>2} \varphi(u)\, \widetilde{\Delta}_n(u)\, du \to 0$ in probability for any test function φ with compact support (see [9]).

Remark 3. A similar result about convergence to a generalized Gaussian process for eigenvalues of random matrices in the Gaussian Unitary Ensemble (GUE) was obtained by Johansson [10] (see further discussion of these results in [9]).

However, a "localized" version of the central limit theorem for random partitions (i.e., for fluctuations at a given point) has not been known as yet. On the one hand, the existence of such a theorem would have seemed quite natural (at least, in the bulk of the partition "spectrum"[3], i.e., for $\lambda_i \in \lambda \vdash n$ such that $i/n \sim x \in (0, 2)$); on the other hand, Kerov's result on generalized convergence might cast some doubt on the validity of the usual convergence.

Note that the asymptotic behavior of fluctuations at the upper edge of the limiting spectrum (corresponding to $x = 0$) is different from Gaussian. As was shown in [2] for λ_1 and in [4, 11, 14] for any λ_i with fixed $i = 1, 2, \dots$,

$$\lim_{n\to\infty} P_n\left\{ \frac{\lambda_i - 2\sqrt{n}}{n^{1/6}} \le z \right\} = F_i(z), \qquad z \in \mathbb{R}, \tag{11}$$

where F_i is the distribution function of the i-th largest point in the so-called Airy random point process, discovered earlier in connection with the limit distribution of the largest eigenvalues for random matrices from the GUE (see [17]). In particular, $F_1(\cdot)$ is known as the Tracy–Widom distribution function.

From the point of view of Kerov's limit theorem (see (10)), the extreme values $\lambda_1, \lambda_2, \dots$ might present a danger, since according to formula (11) the fluctuations of the process $\Delta_n(x)$ near $x = 0$ are large (of order of $n^{1/6}$). As this theorem shows, the edge of the spectrum in fact does not give any considerable contribution into the integral fluctuations. Let us stress, however, that the situation in the bulk of the spectrum remained unclear.

2. MAIN RESULTS

In our first result, we establish the central limit theorem for the random variable $\Delta_n(x)$ given by (6). Set

$$Y_n(x) := \frac{2\theta_x \Delta_n(x)}{\sqrt{\log n}}\,, \tag{12}$$

[3] We use the term "spectrum" informally by analogy with the GUE, to refer to the variety of partition's terms $\lambda_i \in \lambda$ (cf. the book [1] where this term is used in a general context of combinatorial structures characterized by their components).

where $\theta_x = \arccos \frac{\omega(x)-x}{2}$ is the value of the parameter θ in equations (4) corresponding to the coordinates x and $y = \omega(x)$.

Theorem 1. *For each $0 < x < 2$, the distribution of the random variable $Y_n(x)$ with respect to the Plancherel measure P_n converges, as $n \to \infty$, to the standard normal distribution $\mathcal{N}(0,1)$.*

Remark 4. One can show that Theorem 1 also holds for $Y_n(x_n)$ if $x_n \to x \in (0,2)$ as $n \to \infty$.

The local structure of correlations of the random process $\Delta_n(x)$ is described by the following theorem. We write $c_n \asymp 1$ if $c_n n^{\varepsilon} \to \infty$, $c_n n^{-\varepsilon} \to 0$ for any $\varepsilon > 0$, and $a_n \asymp b_n$ if $a_n/b_n \asymp 1$.

Theorem 2. *Fix $x_0 \in (0,2)$ and let $x_1, \dots, x_m \in (0,2)$ be such that $|x_0 - x_i| \asymp n^{-s_i/2}$, where $0 \le s_i \le 1$ $(i = 1, \dots, m)$. For $i = 0$, set formally $s_0 = 1$. Then the random vector $(Y_n(x_0), \dots, Y_n(x_m))$ converges in distribution, as $n \to \infty$, to a Gaussian vector $(Z_{s_0}, \dots, Z_{s_m})$ with zero mean and covariance matrix K with the elements $K(s_i, s_i) = 1$, $K(s_i, s_j) = \min\{s_i, s_j\}$ $(i \ne j)$.*

Note that by Theorem 2, the covariance between $Y_n(x)$ and $Y_n(x')$ asymptotically decays as the distance $|x - x'|$ grows:

$$|x - x'| \asymp n^{-s/2} \quad \Rightarrow \quad \lim_{n\to\infty} \operatorname{Cov}\bigl(Y_n(x), Y_n(x')\bigr) = s.$$

In particular, if $|x - x'| \asymp n^{-1/2}$ (i.e., $s = 1$), then $(Y_n(x), Y_n(x')) \xrightarrow{d} (Z_1, Z_1)$, while if x' is at a fixed distance from x (i.e., $s = 0$) then $Y_n(x)$, $Y_n(x')$ are asymptotically independent.

Remark 5. Results similar to Theorems 1 and 2 were obtained by Gustavsson [8] for eigenvalues in the bulk of the spectrum of random matrices in the GUE.

Let us point out that Theorems 1 and 2 can be reformulated in coordinates u, v (see Sect. 1). To this end, one needs to find the sliding projection (divided by $\sqrt{2}$) of the deviation $\widetilde{\Delta}_n(u)$ (see (7)) onto the line $u + v = 0$ along the tangent of the graph $\Omega(\cdot)$ at point u. Differentiating equations (4), we get

$$\frac{dv}{du} = \frac{x'_\theta + y'_\theta}{x'_\theta - y'_\theta} = \frac{2\theta_x}{\pi} - 1\,,$$

which implies

$$\widetilde{\Delta}_n(u) = \frac{2\theta_x}{\pi}\,\Delta_n(x_n)\,(1 + \eta_n)\,,$$

where $x_n \to x$ and $\eta_n \to 0$ (in probability). Hence, setting

$$\widetilde{Y}_n(u) := \frac{\pi \widetilde{\Delta}_n(u)}{\sqrt{\log n}}\,, \qquad u \in \mathbb{R}, \tag{13}$$

and using Remark 4, we obtain the following elegant versions of Theorems 1 and 2, where the normalization constant is the same for all points.

Theorem 1′. *For each $-2 < u < 2$, the distribution of the random variable $\widetilde{Y}_n(u)$ with respect to the Plancherel measure P_n converges, as $n \to \infty$, to the normal distribution $\mathcal{N}(0,1)$.*

Theorem 2′. *Let $u_0, \dots, u_m \in (-2,2)$, $|u_0 - u_i| \asymp n^{-s_i/2}$, $0 \le s_i \le 1$, with the conventions as in Theorem 2. Then the random vector $\bigl(\widetilde{Y}_n(u_0), \dots, \widetilde{Y}_n(u_m)\bigr)$ converges in distribution, as $n \to \infty$, to a Gaussian vector $(Z_{s_0}, \dots, Z_{s_m})$ with zero mean and the same covariance matrix K.*

Remark 6. The covariance function $K(s,s')$ of Theorems 2 and 2′ determines a Gaussian process Z_s on $[0,1]$ (with zero mean), which can be represented as

$$Z_s \stackrel{d}{=} W_s + \zeta_s \sqrt{1-s}\,, \qquad 0 \le s \le 1,$$

where W_s is a standard Wiener process and $\{\zeta_s,\ 0 \le s \le 1\}$ is a family of mutually independent random variables (also independent of W_s) with normal distribution $\mathcal{N}(0,1)$. This decomposition shows that the process Z_s is highly irregular (e.g., stochastically discontinuous everywhere except at $s=1$), which is a manifestation of asymptotically fast oscillations of the process $\Delta_n(x)$ (as well as $\widetilde{\Delta}_n(u)$) in the vicinity of each point $x \in (0,2)$ (respectively, $u \in (-2,2)$).

In conclusion of this section, let us comment on the asymptotics of the random function $\Delta_n(x)$ at the ends of the limit spectrum, that is, for $x = 0$ and $x = 2$. By the definition (3), $\Delta_n(0) = \lambda(0) - \sqrt{n}\,\omega(0) = \lambda_1 - 2\sqrt{n}$, and according to (11)

$$\lim_{n\to\infty} P_n \left\{ \frac{\Delta_n(0)}{n^{1/6}} \le z \right\} = F_1(z),$$

where $F_1(\cdot)$ is the Tracy–Widom distribution (see Sect. 1). However, the limit distribution of $\Delta_n(2) = \lambda(2\sqrt{n}\,)$ proves to be discrete.

Theorem 3. *Under the Plancherel measure P_n, for any $z \ge 0$*

$$\lim_{n\to\infty} P_n \bigl\{ \Delta_n(2) \le z \bigr\} = F_i(0), \qquad i = [z] + 1,$$

where $F_i(\cdot)$ is the distribution function of the i-th largest point in the Airy ensemble (*see* (*11*)).

Indeed, using the invariance of the measure P_n under the transposition $\lambda \leftrightarrow \lambda'$ (see Sect. 1), we have, due to (11),

$$P_n\bigl\{\Delta_n(2) \le z\bigr\} = P_n\bigl\{\lambda_i' < 2\sqrt{n}\bigr\} = P_n\left\{\frac{\lambda_i - 2\sqrt{n}}{n^{1/6}} < 0\right\} \to F_i(0).$$

Remark 7. In the "rotated" coordinates u, v, a similar result holds for both edges:

$$\lim_{n\to\infty} P_n\bigl\{\tfrac{1}{2}\widetilde{\Delta}_n(\pm 2) \le z\bigr\} = F_i(0), \qquad i = [z] + 1.$$

3. POISSONIZATION

The proof of Theorems 1 and 2 is based on a standard poissonization technique (see, e.g., [2]). Let $\mathcal{P} = \cup_{n=0}^{\infty} \mathcal{P}_n$ be the set of partitions of all natural numbers (as usual, it is convenient to include here the case $n = 0$, where there is just one, "empty" partition $\lambda_\emptyset \vdash 0$). Set $|\lambda| := \sum_{\lambda_i \in \lambda} \lambda_i$ and for $t > 0$ define the poissonization P^t of the measure P_n as follows:

$$P^t(\lambda) := e^{-t}\, t^{|\lambda|} \left(\frac{d_\lambda}{|\lambda|!}\right)^2, \qquad \lambda \in \mathcal{P}. \tag{14}$$

Formula (14) defines a probability measure on the set $\mathcal{P}$, since for $\lambda \in \mathcal{P}_n$ we have $|\lambda| = n$ and hence

$$\sum_{\lambda\in\mathcal{P}} P^t(\lambda) = e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} \sum_{\lambda\in\mathcal{P}_n} \frac{d_\lambda^2}{n!} = e^{-t} \sum_{n=0}^{\infty} \frac{t^n}{n!} = 1.$$

We first prove the "poissonized" versions of Theorems 1 and 2. Let $Y_t(x)$ be given by formula (12) with n replaced by t.

Theorem 4. *For each $0 < x < 2$, the distribution of the random variable $Y_t(x)$ with respect to the measure P^t converges, as $t \to \infty$, to the standard normal distribution $\mathcal{N}(0, 1)$.*

Theorem 5. *In the notations of Theorem 2, the random vector $(Y_t(x_0), \dots, Y_t(x_m))$ converges in distribution (with respect to the measure P^t) to a Gaussian vector $(Z_{s_0}, \dots, Z_{s_m})$ with zero mean and the same covariance matrix K.*

In order to derive Theorems 1 and 2 from Theorems 4 and 5, respectively, one can use a standard de-poissonization method. According to formula (14), the expression for P^t can be viewed as the expectation of the random measure P_N, where N is a Poisson random variable with parameter t:

$$P^t(A) = E\,[P_N(A)] = e^{-t} \sum_{k=0}^{\infty} \frac{t^k}{k!}\, P_k(A). \tag{15}$$

Since N has mean t and standard deviation $\sqrt{t}$, equation (15) suggests that the asymptotics of the probability $P_n(A)$ as $n \to \infty$ can be recovered from that of $P^t(A)$ as $t \sim n \to \infty$. More precisely, one can prove that

$$P_n(A) \sim P^t(A), \qquad t \sim n \to \infty,$$

provided that variations of the probability $P_k(A)$ are small in the zone $|k-n| \le \mathrm{const}\sqrt{n}$. In the context of random partitions, such a de-poissonization lemma was obtained by Johansson (see [2]).

4. SKETCH OF THE PROOF OF THEOREM 4

Note that, in view of (4), the statement of Theorem 4 is equivalent to saying that for any $z \in \mathbb{R}$

$$\lim_{t\to\infty} P^t \left\{ \lambda(\sqrt{t}x) - \lceil \sqrt{t}x \rceil \le 2\sqrt{t}\cos\theta_x + z\sqrt{\log t} \right\} = \Phi(2\theta_x z), \tag{16}$$

where $\Phi(\cdot)$ is the distribution function of the normal law $\mathcal{N}(0,1)$. Using the Frobenius coordinates $\lambda_i - i$, set

$$\mathcal{D}(\lambda) := \cup_{i=1}^{\infty} \{\lambda_i - i\}, \qquad \lambda \in \mathcal{P}.$$

Consider the semi-infinite interval $I_t := [2\sqrt{t}\cos\theta_x + z\sqrt{\log t}, \infty)$ and let $\#I_t$ be the number of points $\lambda_i - i \in \mathcal{D}(\lambda)$ contained in I_t. Using that the sequence $\lambda_i - i$ is strictly decreasing and recalling the definition (2) of the function $\lambda(\cdot)$, it is easy to see that relation (16) is reduced to

$$\lim_{t\to\infty} P^t\{\lambda \in \mathcal{P} : \#I_t \le \lceil \sqrt{t}x \rceil\} = \Phi(2\theta_x z). \tag{17}$$

For $k = 1, 2, \dots$, define the k-point correlation functions by

$$\rho_k^t(x_1, \dots, x_k) := P^t\big\{\lambda \in \mathcal{P} : x_1, \dots, x_k \in \mathcal{D}(\lambda)\big\} \qquad (x_i \in \mathbb{Z},\ \ x_i \ne x_j).$$

The key fact is that the correlation functions ρ_k^t have a determinantal structure (see [4, 11]):

$$\rho_k^t(x_1, \dots, x_k) = \det[J(x_i, x_j; t)]_{1 \le i,j \le k}\,,$$

with the kernel J of the form

$$J(x,y;t) = \begin{cases} \sqrt{t}\, \dfrac{J_x J_{y+1} - J_{x+1} J_y}{x-y}\,, & x \ne y, \\[2mm] \sqrt{t}\left(J'_x J_{x+1} - J'_{x+1} J_x\right), & x = y, \end{cases}$$

where $J_m = J_m\big(2\sqrt{t}\big)$ is the Bessel function of integral order m.

In this situation, one can apply Soshnikov's theorem [15] (generalizing an earlier result by Costin and Lebowitz [5]), stating that the random variable $\#I_t$ satisfies the central limit theorem:

$$\frac{\#I_t - E[\#I_t]}{\sqrt{\mathrm{Var}[\#I_t]}} \xrightarrow{d} \mathcal{N}(0,1) \qquad (t \to \infty), \tag{18}$$

provided that $\mathrm{Var}[\#I_t] \to \infty$. Thus, in order to derive (17) from (18), it remains to obtain the asymptotics of the first two moments of the random variable $\#I_t$. The next lemma is the main technical (and most difficult) part of the work.

Lemma 1. *Let E^t and Var^t denote expectation and variance, respectively, under the measure P^t. Then, as $t \to \infty$,*

$$E^t\big[\#I_t\big] = \sqrt{t}x - \frac{z\theta_x}{\pi}\sqrt{\log t} + O(1),$$
$$\mathrm{Var}^t\big[\#I_t\big] = \frac{\log t}{4\pi^2}\,(1 + o(1)),$$

The proof of Lemma 1 is based on a direct asymptotic analysis of the expressions for the expectation and variance. In so doing, the calculations are quite laborious and heavily use the asymptotics of the Bessel function $J_m\big(2\sqrt{t}\big)$ in various regions of variation of the parameters.

Finally, note that the proof of Theorem 5 follows similar ideas using Soshnikov's central limit theorem for linear statistics of the form $\sum_i \alpha_i \#I_{t_i}$ [16].

5. LINK WITH KEROV'S RESULT

Let us comment on the link between our results and the limit theorem by Kerov [12] (see Sect. 1). In particular, our goal is to explain heuristically the mechanism of the effects that take place for the process $\widetilde{\Delta}_n$.

Note that if $|u - u'| = n^{-s/2}$, $0 \le s \le 1$, then

$$s = \frac{-2\log|u - u'|}{\log n}\,.$$

That is to say, "time" s indexing the components of the limit Gaussian vector in Theorem $2'$, has the meaning of the logarithmic distance between the points u and u', normalized by $\log n$. From this point of view, Theorem $2'$ implies that

$$\mathrm{Cov}\big(\widetilde{Y}_n(u), \widetilde{Y}_n(u')\big) \approx s_n(u,u') := \min\left\{\frac{-2\log|u-u'|}{\log n}\,, 1\right\}.$$

In fact, in the course of the proof of Theorems 2 and 2′ we obtain that for any $\varepsilon > 0$ there exist constants $C_1, C_2 > 0$ such that for sufficiently large n, the following estimate holds uniformly in $u, u' \in (-2, 2)$:

$$\operatorname{Cov}\bigl(\widetilde{\Delta}_n(u), \widetilde{\Delta}_n(u')\bigr) \le \begin{cases} -C_1 \log|u-u'|, & |u-u'| \le \varepsilon, \\ C_2, & |u-u'| \ge \varepsilon. \end{cases} \tag{19}$$

Consider now the integral of $\widetilde{\Delta}_n(u)$ with respect to a test function φ:

$$\widetilde{\Delta}_n[\varphi] := \int_{-2}^{2} \widetilde{\Delta}_n(u)\,\varphi(u)\,du.$$

Using (19) we have

$$\begin{aligned} \operatorname{Var}\bigl(\widetilde{\Delta}_n[\varphi]\bigr) &= \int_{-2}^{2}\int_{-2}^{2} \varphi(u)\,\varphi(u') \operatorname{Cov}(\widetilde{\Delta}_n(u), \widetilde{\Delta}_n(u'))\,du\,du' \\ &\le -C_1 \iint_{|u-u'|\le\varepsilon} \varphi(u)\,\varphi(u') \log|u-u'|\,du\,du' \\ &\quad + C_2 \iint_{|u-u'|\ge\varepsilon} \varphi(u)\,\varphi(u')\,du\,du' < \infty, \end{aligned}$$

since the function $\log|u-u'|$ is integrable at zero. Therefore, $\widetilde{\Delta}_n[\varphi]$ is bounded in distribution as $n \to \infty$, which helps understand why Kerov's result holds without any normalization (see Sect. 1).

Remark 8. We believe that by sharpening the asymptotic estimates (19), it may be feasible to compute the limit of the variance $\operatorname{Var}(\widetilde{\Delta}_n[\varphi])$ and thus recover the Kerov theorem directly from the analysis of the correlation structure. We will address this issue elsewhere.

Conversely, the limiting process $\widetilde{\Delta}(u)$ defined in (8) can be used to get the information contained in Theorems 1′ and 2′ (at least heuristically). To this end, observe that since the number of terms in a typical partition is close to $2\sqrt{n}$ (see Remark 1), it is reasonable to think that the number of "degrees of freedom" of a random partition $\lambda \vdash n$ is of order of $m \asymp \sqrt{n}$, and hence the random variable $\widetilde{\Delta}_n(u)$, $u \in (-2, 2)$, may be represented by the partial sum of the series (8)

$$S_m(u)|_{u=2\cos\theta} := \frac{2}{\pi} \sum_{k=2}^{m} \frac{X_k \sin(k\theta)}{\sqrt{k}}\,, \qquad \theta \in (0, \pi).$$

Note that for any $u = 2\cos\theta$, $u' = 2\cos\theta'$

$$\begin{aligned} \operatorname{Cov}\bigl(S_m(u), S_m(u')\bigr) &= \frac{4}{\pi^2} \sum_{k=2}^{m} \frac{\sin(k\theta)\sin(k\theta')}{k} \\ &= \frac{2}{\pi^2} \sum_{k=2}^{m} \frac{\cos(k(\theta-\theta'))}{k} - \frac{2}{\pi^2} \sum_{k=2}^{m} \frac{\cos(k(\theta+\theta'))}{k}\,. \end{aligned} \tag{20}$$

The second sum in (20) converges for all $\theta, \theta' \in (0, \pi)$. For $\theta = \theta'$, from (20) we get

$$\operatorname{Var}[S_m(u)] \sim \frac{2}{\pi^2} \sum_{k=2}^{m} \frac{1}{k} \sim \frac{2 \log m}{\pi^2} \sim \frac{\log n}{\pi^2} \,,$$

and it follows that

$$\frac{\pi S_m(u)}{\sqrt{\log n}} \xrightarrow{d} \mathcal{N}(0,1) \qquad (n \to \infty),$$

which is in agreement with Theorem 1′. Moreover, if $u' - u \asymp n^{-s/2}$ (and hence $\theta' - \theta \asymp n^{-s/2}$) then the first sum in (20) is approximated by the integral

$$\int_2^m \frac{\cos(x(\theta' - \theta))}{x}\, dx = \int_{2|\theta'-\theta|}^{m|\theta'-\theta|} \frac{\cos y}{y}\, dy \sim \int_{2|\theta'-\theta|}^{\varepsilon} \frac{1}{y}\, dy$$
$$\sim -\log|\theta' - \theta| \sim \frac{s}{2} \log n.$$

Hence

$$\operatorname{Cov}\bigl(S_m(u), S_m(u')\bigr) \sim \frac{s \log n}{\pi^2} \qquad (n \to \infty),$$

as predicted by Theorem 2′.

Remark 9. As already mentioned (see Remarks 3 and 5 and also a comment after formula (11)), there is similarity between the asymptotic properties of the spectra of random partitions and random matrices from the GUE. Our discussion suggests that the relationship between the generalized type convergence [10] and the localized central limit theorem [8] in the GUE can also be explained using the correlation structure of the spectrum. One can expect that similar relationship may be in place for other classes of random matrices and for more general determinantal random ensembles, but this issue needs to be studied further.

Remark 10. Let $\widetilde{Y}(u)$, $u \in [-2, 2]$, be a random process with independent values, such that $\widetilde{Y}(u)$ has a standard normal distribution for each $u \in (-2, 2)$, and $\widetilde{Y}(\pm 2) = 0$ (a.s.). Our results (see Theorem 2′ and Remark 7 after Theorem 3) imply that the random process $\widetilde{Y}_n(\cdot)$ (see (13)) converges to $\widetilde{Y}(\cdot)$ in the sense of finite dimensional distributions. A natural question may arise as to whether this can be extended to weak convergence. However, it is easy to see that the answer is negative, at least under the natural choice of the space of continuous functions $C[-2, 2]$, because the necessary condition of tightness breaks down (see [3]). Indeed, for any $\delta > 0$, $\varepsilon > 0$ and all $u, u' \in (-2, 2)$ such that $|u - u'| \le \delta$, we have

$$\lim_{n\to\infty} P_n\bigl\{|\widetilde{Y}_n(u) - \widetilde{Y}_n(u')| \ge \varepsilon\bigr\} = P\bigl\{|\widetilde{Y}(u) - \widetilde{Y}(u')| \ge \varepsilon\bigr\} > 0.$$

Analogous remark applies to the process $Y_n(x)$, $x \in [0, 2]$, considered in the space $D[0, 2]$ of right continuous functions with left limits.

ACKNOWLEDGMENTS

This work was done when Z.G. Su was visiting the University of Leeds (United Kingdom) under a Royal Society International Fellowship financially supported by the K.C. Wong Educational Foundation. His research was also partially supported by the National Science Funds of China, Grant No. 10371109. The authors are grateful to A. Okounkov and Ya.G. Sinai for their interest in this work and also to A.V. Gnedin, J.T. Kent and Yu.V. Yakubovich for the discussions and helpful remarks.

REFERENCES

1. R. Arratia, A.D. Barbour, and S. Tavaré. *Logarithmic Combinatorial Structures: A Probabilistic Approach*. European Math. Soc. Publ., Zürich, 2003.

2. J. Baik, P. Deift, and K. Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.* **12** (4), 1119–1178 (1999).

3. P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.

4. A. Borodin, A. Okounkov, and G. Olshanski. Asymptotics of Plancherel measures for symmetric groups. *J. Amer. Math. Soc.* **13** (3), 481–515 (2000).

5. O. Costin and J.L. Lebowitz. Gaussian fluctuation in random matrices. *Phys. Rev. Lett.* **75** (1), 69–72 (1995).

6. P. Deift. Integrable systems and combinatorial theory. *Notices Amer. Math. Soc.* **47** (6), 631–640 (2000).

7. W. Fulton. *Young Diagrams: With Applications to Representation Theory and Geometry*. Cambridge Univ. Press, Cambridge, 1997.

8. J. Gustavsson. Gaussian fluctuations of eigenvalues in the GUE. *Ann. Inst. H. Poincaré Prob. Stat.* **41** (2), 151–178 (2005).

9. V. Ivanov and G. Olshanski. Kerov's central limit theorem for the Plancherel measure on Young diagrams. In: *Symmetric Functions 2001: Surveys of Developments and Perspectives*, S. Fomin, ed. Kluwer, Dordrecht, 2002, pp. 93–151.

10. K. Johansson. On fluctuations of eigenvalues of random hermitian matrices. *Duke Math. J.* **91** (1), 151–204 (1998).

11. K. Johansson. Discrete orthogonal polynomials ensembles and the Plancherel measure. *Ann. Math.* **153** (1), 259–296 (2001).

12. S. Kerov. Gaussian limit for the Plancherel measure of the symmetric group. *C. R. Acad. Sci. Paris, Sér. I Math.* **316** (4), 303–308 (1993).

13. B.F. Logan and L.A. Shepp. A variational problem for random Young tableaux. *Adv. Math.* **26** (2), 206–222 (1977).

14. A. Okounkov. Random matrices and random permutations. *Internat. Math. Res. Notices* **2000** (20), 1043–1095 (2000).

15. A.B. Soshnikov. Gaussian fluctuation for the number of particles in Airy, Bessel, sine, and other determinantal random point fields. *J. Statist. Phys.* **100** (3–4), 491–522 (2000).

16. A.B. Soshnikov. Gaussian limit for determinantal random point fields. *Ann. Probab.* **30** (1), 171–187 (2002).

17. C.A. Tracy and H. Widom. Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159** (1), 151–174 (1994).

18. A.M. Vershik and S.V. Kerov. Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tables. *Dokl. Akad. Nauk SSSR* **233** (6), 1024–1027 (1977) [Soviet Math. Doklady **18** (2), 527–531 (1977)].

19. A.M. Vershik and S.V. Kerov. Asymptotic of the largest and the typical dimensions of irreducible representations of a symmetric group. *Funkts. Anal. i ego Prilozh.* **19** (1), 25–36 (1985) [Func. Anal. Appl. **19** (1), 21–31 (1985)].